

## Neural correlates of acquired color category effects

Alexandra Clifford<sup>a,\*</sup>, Anna Franklin<sup>a</sup>, Amanda Holmes<sup>b</sup>, Vicky G. Drivonikou<sup>a</sup>, Emre Özgen<sup>c</sup>, Ian R.L. Davies<sup>a</sup>

<sup>a</sup> School of Psychology, University of Surrey, UK

<sup>b</sup> Department of Psychology, Roehampton University, London, UK

<sup>c</sup> Department of Psychology, Bilkent University, Turkey

### ARTICLE INFO

#### Article history:

Accepted 30 April 2012

Available online 19 June 2012

#### Keywords:

Color categories

Category training

Color

Event-Related Potentials

### ABSTRACT

Category training can induce category effects, whereby color discrimination of stimuli spanning a newly learned category boundary is enhanced relative to equivalently spaced stimuli from within the newly learned category (e.g., categorical perception). However, the underlying mechanisms of these acquired category effects are not fully understood. In the current study, Event-Related Potentials (ERPs) were recorded during a visual oddball task where standard and deviant colored stimuli from the same or different novel categories were presented. ERPs were recorded for a test group who were trained on these novel categories, and for an untrained control group. Category effects were only found for the test group on the trained region of color space, and only occurred during post-perceptual stages of processing. These findings provide new evidence for the involvement of cognitive mechanisms in acquired category effects and suggest that category effects of this kind can exist independent of early perceptual processes.

Crown Copyright © 2012 Published by Elsevier Inc. All rights reserved.

### 1. Introduction

There is converging evidence that the way in which we see the world is influenced by our concepts and categories (e.g., Goldstone, Gerganov, Landy, & Roberts, 2008; Kay & Kempton, 1984; Lupyan, Thompson-Schill, & Swingle, 2010). For example, cross-linguistic differences in the expression of categories or concepts can lead to corresponding differences in performance on 'perceptual' tasks such as visual search or same-different judgment tasks. These effects are found across a wide range of domains (e.g., Boroditsky, 2003). There are also multiple studies showing that learning novel categories can lead to changes in performance on these types of tasks (e.g., Drivonikou, Clifford, Franklin, & Davies, 2011; Goldstone, 1994; Notman, Sowden, & Özgen, 2005; Özgen & Davies, 2002). For example, training participants to group exemplars into novel categories leads to an increase in speed and/or accuracy in performance when stimuli are from different categories (acquired distinctiveness) and/or a decrease in speed and/or accuracy when they are from the same novel category (acquired equivalence). Effects like these demonstrate that category effects such as 'categorical perception' (heightened discrimination of stimuli from different categories, than stimuli from the same category) can be acquired for a range of stimulus domains.

\* Corresponding author. Address: School of Psychology, University of Surrey, Guildford, Surrey GU2 7XH, UK.

E-mail address: [a.clifford@surrey.ac.uk](mailto:a.clifford@surrey.ac.uk) (A. Clifford).

One important question is the extent to which categories and concepts penetrate perception. Tasks such as visual search, same-different judgment or X-AB tasks may reflect perceptual sensitivity, although these tasks are also potentially influenced by higher order mechanisms such as central attention and memory. For example, tasks that involve a delay between stimulus presentation and participant judgments, such as same-different judgment or X-AB tasks, are vulnerable to memory processes, and so it is possible that participants rely on a memory trace for the first stimulus when judging whether a second stimulus is the same or different. It is also possible that linguistic strategies enhance performance on a same-different judgment or X-AB task, whereby appropriate category labels are accessed when stimuli are presented (e.g., Winawer et al., 2007). Therefore, differences in performance on these tasks may reflect differences or changes in these higher order mechanisms rather than actual perceptual change (Goldstone, 1994). In other words, it is possible that acquiring new categories may alter the way in which we attend to or make decisions about exemplars, but not actually change their perceptual representation (Pylyshyn, 1999). In this sense, although these category effects are often termed 'categorical perception', they may actually not be perceptual and the term 'category effect' or 'categorical responding' may be more appropriate (e.g., Franklin, Pilling, & Davies, 2005; Roberson, Hanley, & Pak, 2009). There is evidence however, that concepts and categorization penetrate early stages of perceptual processing (e.g., Lupyan et al., 2010; Notman et al., 2005; Reber, Stark, & Squire, 1998; Tanaka & Curran, 2001). For the case of acquired categorical perception, it appears that newly trained

categories could affect early stages of the perceptual processing stream, perhaps as early as V1 (Notman et al., 2005). Notman et al. (2005) found that following training on categories of novel grating patterns, the resulting acquired category effect was orientation specific. This specificity of the acquired category effect to the trained orientation potentially localizes the effect to orientation selective areas of visual cortex.

Studies of color categories have also provided evidence of category effects early on in perceptual processing. Event-Related Potential (ERP) studies have found category effects for post-perceptual ERP components (e.g., P2 and P3; Holmes, Franklin, Clifford, & Davies, 2009), and also perceptual ERP components occurring as early as 90 ms post-stimulus onset (Clifford, Holmes, Davies, & Franklin, 2010; Holmes et al., 2009) across the blue–green boundary. Additionally, a neuroimaging study (Siok et al., 2009) has identified greater activation for search for targets from a different color category to the distractor (between-category) than the same color category- (within-category), in regions of visual cortex (V2, V3), as well as language related regions of the brain (LH posterior temporoparietal region, middle temporal gyrus and inferior prefrontal cortex). The presence of category effects during early perceptual and post-perceptual stages of processing for blue–green color categories could provide further evidence that categories penetrate and modulate early perception in a top-down manner (Siok et al., 2009). However, for the case of color categories, it is not clear what the origin of these early perceptual category effects is. Infants also respond categorically across the blue–green boundary (Bornstein, Kessen, & Weiskopf, 1976; Clifford, Franklin, Davies, & Holmes, 2009; Franklin & Davies, 2004; Franklin et al., 2005; Franklin, Drivonikou, Bevis, et al., 2008; Franklin, Drivonikou, Clifford, et al., 2008), and it is possible that the early perceptual category effects reflect ‘in-born’ qualities of the perceptual visual system, rather than perceptual change as a result of top-down modulation from higher-order processes (although infant color categories could also be acquired in this way).

Studies where novel color categories are trained and color category effects are acquired have already established that category effects need not depend on ‘in-born tendencies’ (Drivonikou et al., 2011; Özgen & Davies, 2002). For example, Özgen and Davies (2002) found that for participants who were trained to sort colors from the green region of color space into two novel unnamed categories, there were category effects across the newly trained category boundary on a same-different judgment task, whereas there were no category effects for those who did not receive category training. Drivonikou et al. (2011) using the same category training procedure, found that this acquired category effect is lateralized to the left hemisphere, as has previously been found for category effects across the blue–green boundary under some circumstances (e.g., Drivonikou et al., 2007; Gilbert, Regier, Kay, & Ivry, 2006; Roberson, Pak, & Hanley, 2008, but see also Brown, Lindsey, & Guckes, 2011; Witzel & Gegenfurtner, 2011). Additionally, Zhou et al. (2010) have found the left hemisphere advantage can be acquired for within-category stimuli merely by training participants to give within-category stimuli different names. As these learning effects appeared to be left hemisphere lateralized, and this hemisphere is responsible for language, Zhou et al. suggested that post-perceptual linguistic processes may be involved in acquired CP. However, what is not clear is whether the early perceptual category effects found in the ERP and fMRI studies of blue–green color categories are also present for these acquired color category effects. If category effects are found in early perceptual ERP components for acquired color categories, this would provide further evidence that categories can penetrate early stages of visual processing. If early perceptual category effects are not found for acquired categories, this would suggest that category effects, traditionally termed

‘categorical perception’ can actually exist without the involvement of early perceptual processes.

Here we investigate this issue by exploring the time course of acquired color category effects using the Event-Related Potential technique. Novel color categories were learned during a 3 day training phase that was identical to that employed by Özgen and Davies (2002). Test group participants were trained to form two novel green categories, with the new category boundary situated in the center of the green region of color space. After successful completion of the training phase, participants undertook a refresher training session and then completed a visual oddball task similar to that used in Holmes et al.’s (2009) investigation. Visual oddball tasks involve detection of infrequent (deviant) stimuli among high-frequency (standard) stimuli and are particularly appropriate for exploring category effects as the waveforms elicited by deviant stimuli correspond to processes involved in event categorization. The succession of ERP components elicited by stimuli within an attended oddball task such as this is: P1, N1, P2, N2 and P3. The P1 component and the N1 component (occurring ~80–120 ms and ~130–190 ms post-stimulus onset, respectively) reflect early perceptual and sensory processes in the brain (e.g., Polich, 1999). These components are sensitive to the physical characteristics of stimuli but have also been shown to be modifiable by manipulations of attention (e.g., Mishra, Martínez, Schroeder, & Hillyard, 2012; Taylor, 2002). The P2, N2, and P3 components (occurring ~210–270 ms, ~280–340 ms, ~350–600 ms, post-stimulus onset, respectively), primarily correspond to post-perceptual stimulus evaluation processes (Patel & Azzam, 2005), and incorporate essentially any post-perceptual process that is not related to the preparation of an overt response (e.g., McCarthy & Donchin, 1981).

During the visual oddball task used in the current study, ERPs were recorded in response to stimuli from the green region of color space (on which the test group had been trained) and stimuli from the blue region of color space (on which the test group had no prior training). Each experimental block consisted of a standard stimulus that appeared frequently and two deviant stimuli that appeared infrequently. For the test group who had been trained to form two new categories in the green region of color space, each of the green deviants varied in their categorical relationship to the green standard. One of the deviants was from the same new category as the standard (within-category) and one was from the other new category (between-category). As there was no category training for the blue region, both blue deviants were from the same established category as the blue standard. The standard-deviant hue separations were equivalent for blue and green stimulus sets. As an additional control, a separate group of participants who received no category training also completed the visual oddball task (the control group). For these participants the standard and oddball stimuli were all from the same established color category (i.e., blue or green). Therefore each of the deviants had the same categorical relationship to the standard despite being located in different areas of the established category (i.e., near to the center of the established category vs. close to an established category boundary).

It was predicted that category learning would enhance detection of deviants when stimulus pairs crossed a newly learned category boundary, consistent with the findings of Özgen and Davies (2002). Additionally, as found by Özgen and Davies (2002) and Drivonikou et al., 2011, we hypothesized that enhanced detection for deviants located close to an established category boundary would be shown for untrained regions. ERP waveforms for deviant stimuli belonging to a different newly learned category to the standard and deviant stimuli from the same newly learned category as the standard were compared, and the time course of differences noted.

	Day 1	Day 2	Day 3	Test Day
	Context Training	Context Training	Context Training	Refresher Training
<b>Test group</b>	+	+	+	+
	Singleton Training	Singleton Training	Singleton Training	ERP task
<b>Control group</b>	No Training			ERP task

**Fig. 1.** Summary of the experiment schedule. Participants in the test group were trained in a categorization task in three sessions over 3 days. On day four or five participants received a refresher session, followed by the ERP oddball task. Participants in the control group performed the ERP task without any of the training sessions or the refresher.

## 2. Method

### 2.1. Overview

Participants in the test group completed a training phase over three sessions occurring on three consecutive days. During training, participants learned to categorize stimuli from the green region of color space into two new color categories distinguished by a new boundary falling roughly in the center of the green category. Each session of training consisted of 'context training' followed by 'singleton training'. On the test day participants undertook a refresher training session, followed by an ERP task. Participants in the control group performed the ERP task with no prior training. A summary of the experiment schedule can be seen in Fig. 1.

### 2.2. Training phase

The apparatus and experimental set-up and the stimuli were identical for each of the three types of training within the training phase (context training, singleton training and refresher training) but the design and procedure of each differed.

#### 2.2.1. Participants

A group of twenty paid volunteers (9 males, 11 females) with a mean age of 25.1 years ( $SD = 4.9$ ) took part in the training phase, hereafter referred to as the 'test group'. All had normal or corrected-to-normal vision and had no sign of red-green color vision deficiency as assessed by the Ishihara test for color blindness (Ishihara, 1987). Informed consent was obtained from all participants.

#### 2.2.2. Apparatus and experimental set-up

A 21-in. Sony Trinitron CRT monitor (model GDM-F520) with a Dell Pentium 4 computer was used and stimulus presentation was controlled with Visual Basic software (with ExactTics high resolution timer). Participants were seated in a darkened laboratory, approximately 70 cm away from the computer screen and at eye-level to the center of the monitor.

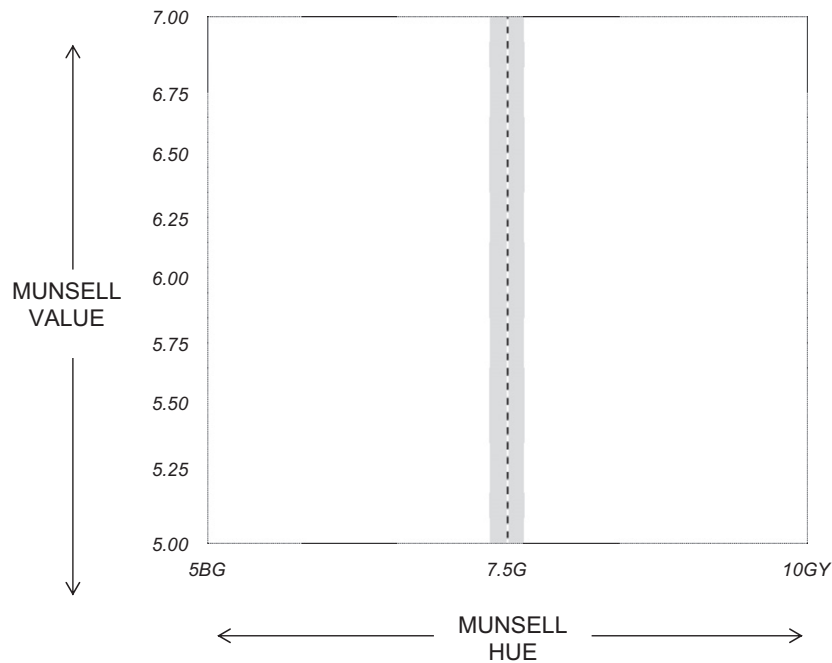
#### 2.2.3. Stimuli

The stimuli were computer-generated colors randomly taken from an area within the green region of color space. Stimuli were taken from the Munsell color system<sup>1</sup> and were emulated on a calibrated monitor by converting Munsell stimuli to  $x$ ,  $y$ ,  $Y$  (CIE, 1931; see Wyszecki & Styles, 1982) color space. The trained hue boundary fell roughly in the center of the green linguistic category (7.5G) and stimuli were generated randomly from the region around this. Munsell Hue varied between 5BG and 10GY and Munsell Value varied between 5 and 7. Munsell Chroma was kept constant at 6. Stimuli within 0.2 Munsell Hue units of the boundary were not used (see Fig. 2). Stimuli were 5 cm colored squares displayed on a neutral gray background ( $Y = 30.00 \text{ cd/m}^2$ ,  $x = 0.313$ ,  $y = 0.319$ ). The chromaticity co-ordinates of stimuli were checked and verified with a Cambridge Research Instruments ColorCal colorimeter at regular intervals during the study.

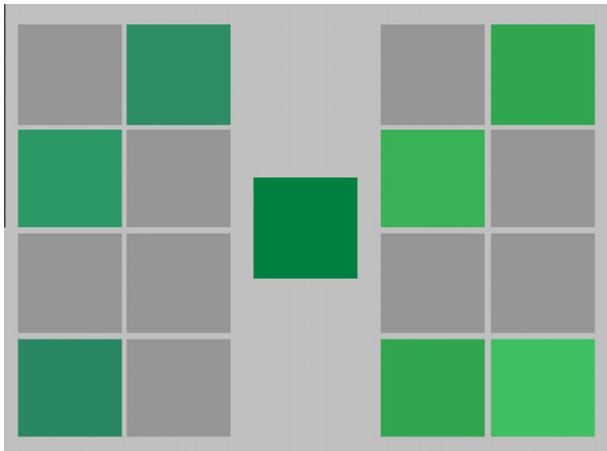
#### 2.2.4. Context training design and procedure

For each training trial, a random point within the green training area was selected and the corresponding color was displayed on the monitor. The random test color could fall on either side of the training boundary, avoiding points very close to it. This color was presented in the center of the screen on a uniform gray background, flanked on either side by two grids of eight slots (two columns, four rows) to be filled with the incoming test colors (see Fig. 3). Participants could place the first test color in either grid. Once the first color was placed on the left or right side, colors from the same category had to be placed on the same side of the screen and colors from the other category had to be placed on the opposite side. Response keys were assigned to select which grid each test color belonged to ('left-arrow' for category 1 or 'right-arrow'

<sup>1</sup> Munsell is a color metric based on an extensive program of psychophysical judgments and it is intended to be standardized (Newhall, Nickerson, & Judd, 1943). There are three dimensions: Hue, Value (lightness) and Chroma (colorfulness, rather like saturation). Stimuli are specified using a notation that consists of three co-ordinates that each relate to one of the three dimensions. The metric has been commonly used in studies of color category effects (e.g., Bornstein & Korda, 1984; Gilbert et al., 2006), although inequalities in the metric have been highlighted (see Brown et al., 2011). The current study transcends these metric issues as category effects were only found after category training. If these effects were due to inequalities in the metric then they should also be present before category training and in the control group who have not received category training.



**Fig. 2.** Random color generation area from which the training stimuli were produced. The dashed line represents the new category boundary that participants learned during training. Points closer than 0.2 Munsell Hue steps from the boundary were never used as stimuli (this excluded area is represented by the shaded region).



**Fig. 3.** A representation of the context training phase. Randomly generated stimuli appeared individually in the center of the screen and participants had to make a category judgment, deciding whether each stimulus belonged to the category on the right or the left of the screen.

for category 2). Test colors remained in the center of the screen until a key press response was made and immediate feedback was given after each response. If the response was correct the color moved to the designated grid and remained in its slot, but if the response was incorrect the color moved to the designated grid and then disappeared, accompanied by a sound indicating the incorrect categorization. Participants were not given instructions on what the categorization was based on but had to learn from the feedback to complete the context training phase. When all sixteen slots were filled with correctly identified test colors a set was complete and a new one began. This first stage of training ended when a minimum of 20 sets were completed with at least three of these being error free. Participants were informed that the task would end when their performance was 'sufficiently good'. A typical participant completed the context training in around 30 min.

#### 2.2.5. Singleton training design and procedure

In this second stage of training, single test colors were presented in the center of the screen on a uniform gray background in the absence of the two grids that were present in the context training. The participants' task was to decide which of the two categories each singleton belonged to, indicating their choice with a key press ('left-arrow' or 'right-arrow'). Each color remained on the screen until a response was made. Feedback was provided throughout the task with incorrect responses signaled by the appearance of the word 'incorrect' on the screen, accompanied by a short sound. Correctly identified colors no longer remained on the screen, so that there was no visual point of reference for each categorization, unlike in the context training. The criteria for completing the singleton training were at least 250 complete trials including 25 consecutive correct responses. Participants were informed that the task would end when their performance was 'sufficiently good' and were usually able to complete it in around 10 min.

#### 2.2.6. Refresher training design and procedure

Participants in the test group undertook refresher training on the test day (day four). The refresher phase consisted of a shorter version of context training and singleton training. Participants were required to complete a minimum of 10 sets of context training with at least one error free set and at least 100 complete trials of singleton training, with 25 consecutive correct responses. On average, context training took around 10 min and singleton training approximately 5 min.

### 2.3. ERP task

#### 2.3.1. Participants

All participants who took part in the training phase (the test group) also completed the ERP task. Six of these participants were excluded from the ERP analysis due to an insufficient number of artifact free trials, leaving a final sample of 14 (7 males, 7 females) with a mean age of 24.1 years ( $SD = 3.2$ ). An additional group of

fourteen participants (7 females and 7 males; mean age = 25.4 ( $SD = 4.5$ )) undertook the ERP task with no prior training, hereafter referred to as the control group. All had normal or corrected-to-normal vision and were also screened for red–green color vision deficiency using the Ishihara test for color blindness (Ishihara, 1987). Data from four additional control group participants were not included due to an insufficient number of artifact free trials (see ‘ERP recording’ for further information on artifact rejection).

### 2.3.2. Apparatus and experimental set-up

The experimental set up was identical to that of the training phase.

### 2.3.3. Stimuli

The stimuli were from the blue–green region of color space and consisted of two sets of four colors (1 blue set; 1 green set; see Fig. 4) varying only in Munsell Hue; Chroma (saturation) and Value (lightness) were both constant at 6 (see Table 1 for the Munsell codes and  $Y, x, y$  (CIE, 1931) chromaticity coordinates of the stimuli). Within each set, stimuli were equidistant with adjacent pairs separated by 5 Munsell Hue units. Stimuli were 9 cm colored squares individually presented in the center of the screen, subtending a maximum visual angle of  $7.3^\circ$ . The colored squares were displayed on a neutral gray background ( $Y = 30.00 \text{ cd/m}^2$ ,  $x = 0.313$ ,  $y = 0.319$ ). Chromaticity co-ordinates of stimuli were verified with a Cambridge Research Instruments ColorCal colorimeter at regular intervals throughout the data collection period.

The two sets of four colors ( $1 \times$  blue set;  $1 \times$  green set) were further divided into four triads of adjacent stimuli. Each triad consisted of a standard plus two deviants, which were equidistant in color space to the standard. Deviants are referred to in terms of their location within the established category structure. This is because in the majority of cases (i.e., green and blue conditions for the control group and the blue condition for the test group), stimuli have no categorical context, as each stimulus in the triad belongs to the same color category. Therefore, one deviant in each triad is referred to as the ‘central’ deviant as it is the deviant located in the center of the established category. The other deviant in each triad is referred to as the ‘outer’ deviant as it is located towards

**Table 1**

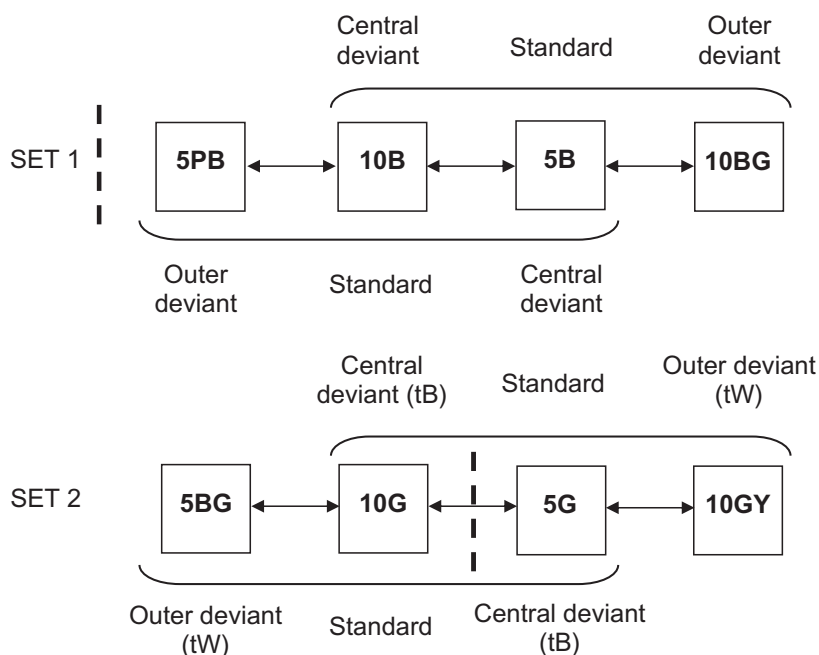
Munsell codes and  $Y, x, y$  (CIE, 1931) chromaticity coordinates of the stimuli. The white point of the monitor was  $Y = 77.76 \text{ cd/m}^2$ ,  $x = 0.326$ ,  $y = 0.341$ .

Munsell code		$Y$	$x$	$y$
5PB 6/6	Blue 1	23.3	0.253	0.256
10B 6/6	Blue 2	23.3	0.239	0.265
5B 6/6	Blue 3	23.3	0.232	0.278
10BG 6/6	Blue 4	23.3	0.233	0.301
5BG 6/6	Green 1	23.3	0.244	0.329
10G 6/6	Green 2	23.3	0.259	0.356
5G 6/6	Green 3	23.3	0.274	0.38
10GY 6/6	Green 4	23.3	0.313	0.418

the outer edge of the established category (see Fig. 4 for a representation of the stimulus triads). However, for the test group, a new category boundary in the center of the existing green category has been established through training and a categorical context for standard and deviant stimuli has thus been provided. For the test group therefore, green deviant stimuli are referred to either as ‘central deviant (tB)’ (where ‘tB’ denotes a deviant that is from a different category to the standard for the test group: test group between-category) or ‘outer deviant (tW)’ (where ‘tW’ denotes a deviant that is from the same category as the standard for the test group: test group within-category). For the sake of brevity, the tB and tW labels are also used when referring to all conditions collectively (i.e., across test and control groups).

### 2.3.4. Design and procedure

There were four blocks of 276 trials with one block for each triad. Within each block the standard was repeated 228 times (82.6%), and outer deviants (tW) and central deviants (tB) each occurred 24 times (8.7%). This resulted in a total of 1104 trials (912 standard; 96 outer deviant (tW); 96 central deviant (tB)). Half of the trials for each stimulus type consisted of stimuli from the green region of color space (Set 2) (456 standard; 48 outer deviant (tW); 48 central deviant (tB)) and half consisted of stimuli from the blue region of color space (Set 1) (456 standard; 48 outer deviant; 48 central deviant). An additional 12 deviant stimuli were included



**Fig. 4.** A representation of the stimulus triads. The dashed lines in Set 1 represent the established category boundaries. The dashed line in Set 2 represents the new category boundary that participants learned during training. The notations in each box refer to Munsell codes and the brackets indicate stimulus triads.



so that the number of deviants would not be identical for each block. As participants' task was to keep count of the number of deviants, this was done to ensure that participants could not anticipate the number of deviant stimuli in each block and were therefore more likely to fully attend to the task. These additional stimuli were distributed in three of the four blocks: one of the blocks contained 3 additional outer deviants (tW), one of the blocks contained 3 additional central deviants (tB) and in the final block, 6 instances of deviant category stimuli (3 outer deviant (tW), and 3 inner deviant (tB)) were integrated. These additional deviant stimuli were not included in the ERP analysis.

Each block was divided into three sessions of 92 trials, and there was a self-paced break in between each session. Stimuli were presented in a random order but with two provisos: each sequence of 92 trials started with the presentation of 8 standard stimuli, and there were no immediate repetitions of deviant stimuli. The order of the 4 blocks was randomized across participants. Each trial began with a central fixation cross, presented on a gray background. The duration of this 'inter-trial-interval' (ITI) varied randomly between 1300 and 1600 ms. Following this, the stimulus display appeared for 400 ms. Participants were asked to mentally count the number of deviant (target) stimuli within each consecutive sequence of 92 trials. A short pause was allowed at the end of each sequence for the participant to report the number of deviant stimuli to the experimenter. Participants were instructed to maintain their gaze at the center of the screen, and were also asked to minimize eye blinks and to avoid any unnecessary movement.

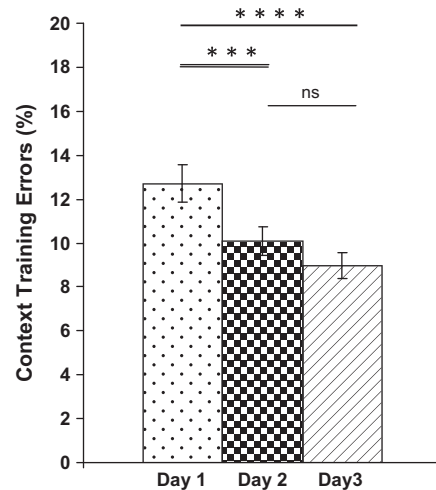
### 2.3.5. ERP recording

EEG was recorded from Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Iz, O2, PO9, PO10 and the average of left and right earlobe references according to the 10–20 system (Jasper, 1958), using Ag–AgCl electrodes. Horizontal electro-oculogram (HEOG) was recorded bipolarly from the outer canthi of each eye. Electrode impedances were kept below 5 k $\Omega$ , and EEG and EOG were sampled on-line with a digitization rate of 1000 Hz and were digitally filtered with a band-pass filter of 0.1–100 Hz using Neuroscan software (version 4.3). Following EEG recording, a low-pass filter of 40 Hz was applied and the data were down-sampled to 200 Hz to save later computation time. EEG and HEOG were epoched off-line with a window extending to 900 ms after stimulus onset, relative to a 100 ms pre-stimulus baseline. A process of artifact rejection excluded trials with lateral eye movements (HEOG exceeding  $\pm 30$   $\mu$ V), as well as trials with vertical eye movements, eye blinks (Fp1/Fp2 exceeding  $\pm 60$   $\mu$ V), or other artifacts (a voltage exceeding  $\pm 60$   $\mu$ V at any electrode) measured after stimulus onset. Criteria for artefact rejection were determined on the basis of previous research (see e.g., Dalvit & Eimer, 2011). A minimum of 24 artifact free trials for each stimulus condition was required for inclusion of any one participant's data in the final sample. This number is comparable to previous studies that have used a similar paradigm to investigate category effects using ERPs (see Holmes et al., 2009). For each participant trial numbers were matched across stimulus conditions to the condition containing the least number of accepted trials. Included trials were selected randomly from all available trials within each condition. The resulting mean number of trials per stimulus condition was 26.5 ( $SD = 2.5$ ).

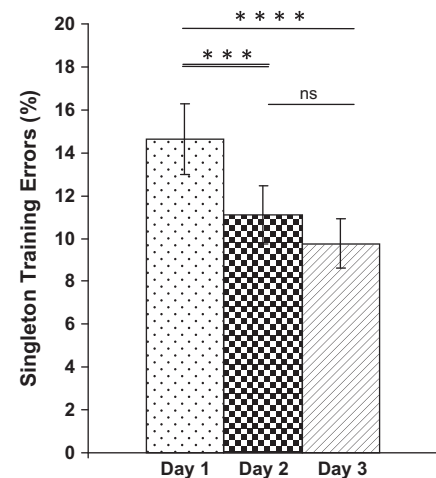
## 3. Results

### 3.1. Training phase

All participants successfully completed the training phase. Participants completed between 500 and 1000 context training



**Fig. 5.** Percentage errors made on the context training phase across the three training days. Significant differences are starred: \*\*\*\* $p < .001$ ; \*\*\* $p < .005$ ; \* $p < .05$ . Error bars represent  $\pm 1SE$ .



**Fig. 6.** Percentage errors made on the singleton training phase across the three training days. Significant differences are starred: \*\*\*\* $p < .001$ ; \*\*\* $p < .005$ ; ns = non-significant. Error bars represent  $\pm 1SE$ .

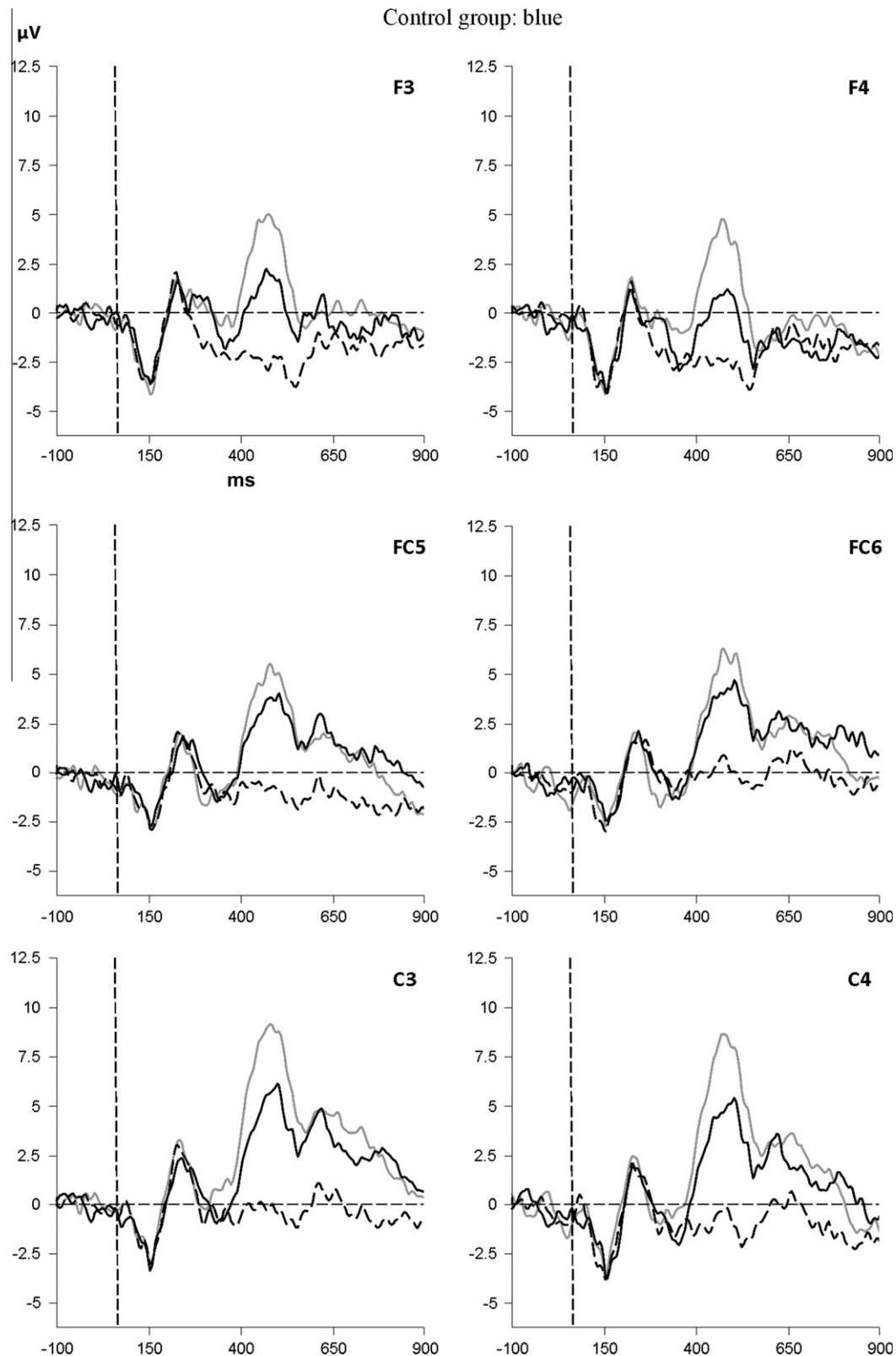
trials and between 250 and 500 singleton color trials on each day. On the test day, participants completed a refresher training session during which they performed between 150 and 250 context trials and between 100 and 250 singleton trials. Therefore, throughout the course of the training phase, the mean number of categorization trials completed by participants was 3300 trials.

#### 3.1.1. Training phase data analysis

To investigate performance across the training phase, the percentage of errors (incorrect classifications) was calculated for context training and singleton training on each of the three training days. Statistical analysis was conducted using Analysis of Variance (ANOVA) to compare the percentage of incorrect responses made on each day, for the context and singleton training phases. Where significant differences were found, paired samples Bonferroni corrected  $t$ -tests were conducted to investigate the pattern of improvement from day to day.

#### 3.1.2. Context training

For the mean percentage errors made on each day see Fig. 5. One-way ANOVA confirmed that there was a significant difference



**Fig. 7.** Control group grand-averaged ERP waveforms elicited in the 900 ms interval following stimulus onset (0 ms) in response to standard (dashed lines), outer deviant (tW; gray lines), and central deviant (tB; black lines) blue color stimuli. The horizontal dashed line represents baseline (0 µV).

in performance on the context training phase across the 3 days ( $F(2,39) = 35.82$ ,  $p < .001$ ,  $\eta_p^2 = .65$ ). To explore the pattern of performance paired samples  $t$ -tests were used to compare the number of errors made on days one and two, days two and three

and days one and three. These revealed that there were significantly fewer errors made on day two compared to day one ( $t(13) = 5.93$ ,  $p < .001$ ,  $d = 0.79$ ). Although improvement continued on day three, performance on days two and three was not

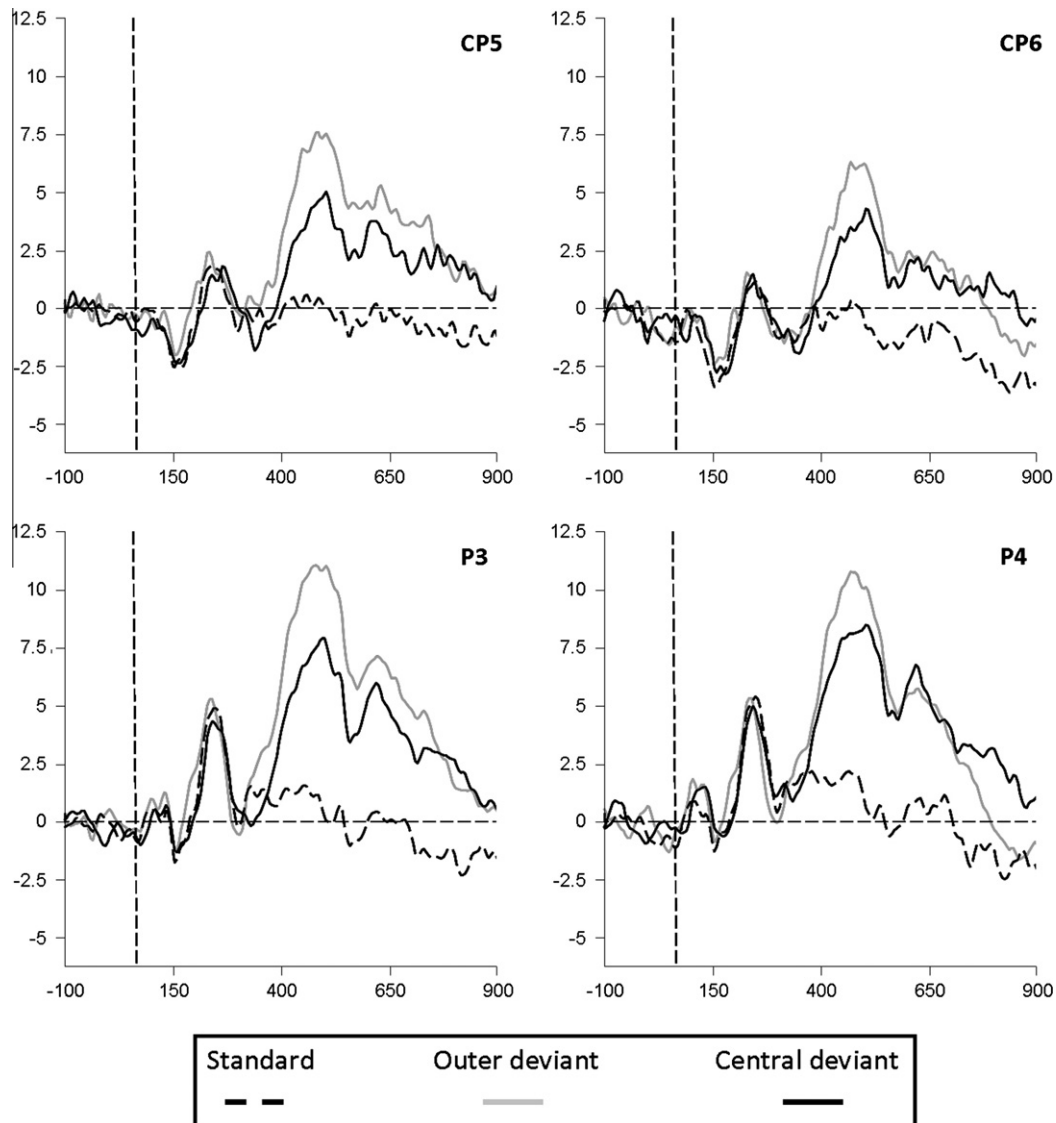


Fig. 7 (continued)

significantly different ( $t(13) = 2.79$ ,  $p = .022$ ,  $d = 0.12$ ). Overall, there was significant improvement between day one and day three ( $t(13) = 7.29$ ,  $p < .001$ ,  $d = 1.19$ ).

### 3.1.3. Singleton training

See Fig. 6 for the mean percentage singleton training errors made on each day. One-way ANOVA revealed that there was a significant difference in the percentage of errors made on singleton training on each of the 3 days ( $F(2,39) = 17.77$ ,  $p < .001$ ,  $\eta_p^2 = .48$ ). To investigate this further paired samples  $t$ -tests were conducted as for the context training analysis. These confirmed that performance on the second day was significantly better than the first day ( $t(13) = 3.73$ ,  $p = .001$ ,  $d = 0.53$ ). However, performance on the third day, although reduced, was not significantly different from day two ( $t(13) = 1.88$ ,  $p = .075$ ,  $d = 0.13$ ). Overall, performance improved significantly from day one to day three ( $t(13) = 5.64$ ,  $p < .001$ ,  $d = 0.79$ ).

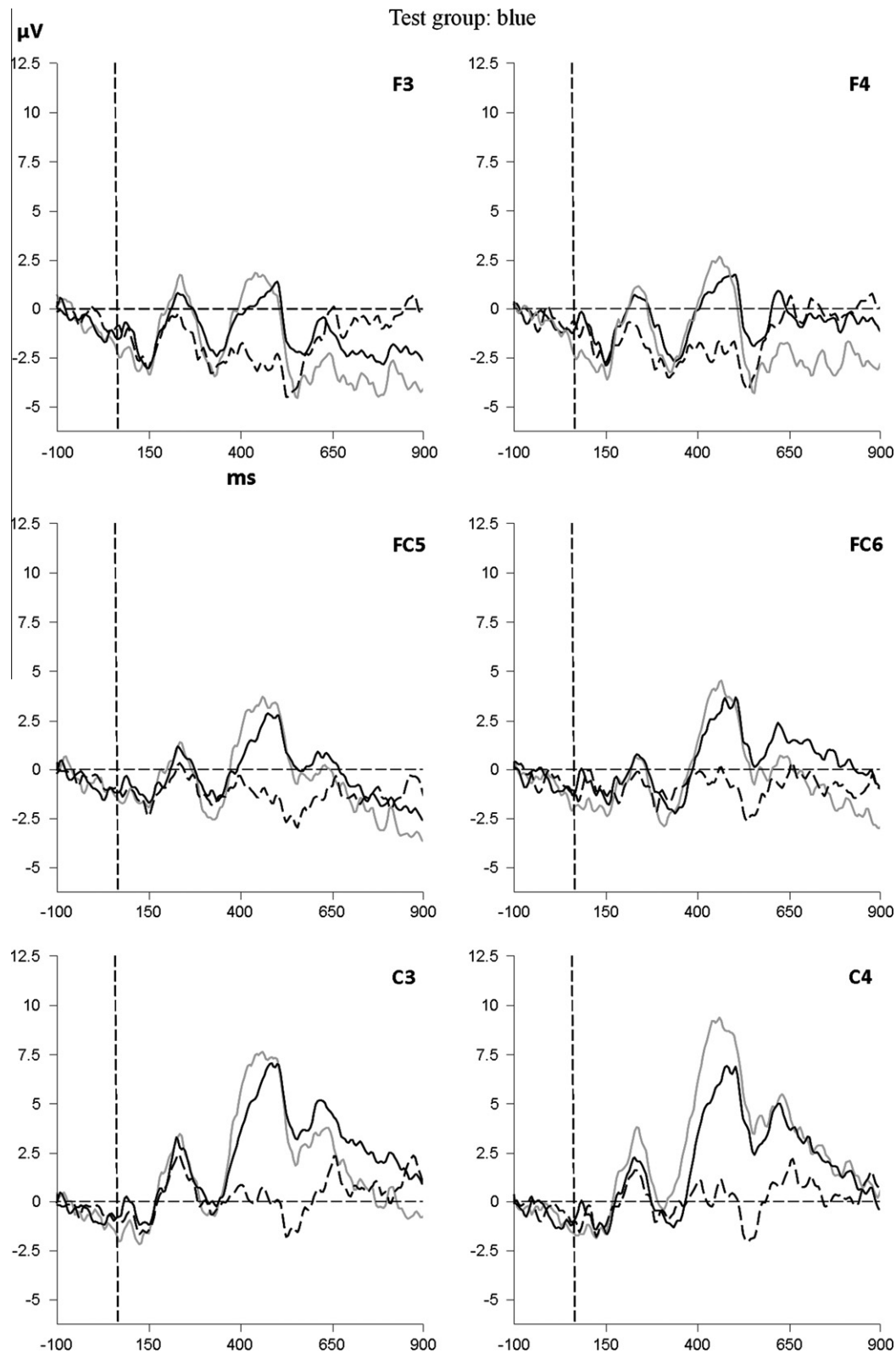
## 3.2. Visual oddball task

### 3.2.1. ERP data analysis

Separate averages were computed for outer deviants (tW) and central deviants (tB). ANOVAs were conducted on ERP mean

amplitudes and peak latencies obtained for specific sets of electrodes within predefined measurement windows. Analyses focused on electrodes at anterior (F3/4, F7/8, FC5/6), posterior (CP5/6, P3/4, P7/8), and midline (Fz, Cz, Pz) scalp regions. ERP mean amplitudes and peak latencies for outer deviants (tW) and central deviants (tB) were calculated for successive post-stimulus time intervals (P1-window: 80–120 ms; N1-window: 130–190 ms; P2-window: 210–270 ms; N2-window: 280–340 ms; P3-window: 350–600 ms). These time windows were determined on the basis of prior research (see Rugg & Coles, 1995, for a review) and inspection of individual subject waveforms. Category effects were tested by comparing the mean amplitudes and peak latencies elicited by outer deviants (tW) and central deviants (tB). For each of the identified time intervals (P1, N1, P2, N2, P3), separate four-way Repeated Measures ANOVAs were performed at each scalp region (anterior, midline, posterior) with Stimulus (outer deviant (tW), central deviant (tB)), Color Region (green, blue) and Electrode as within subjects factors and Group (test, control) as a between subjects factor. The reporting of results will focus on the three-way interaction between Stimulus, Color Region and Group as this crucial interaction reveals whether there has been an impact of training. When this interaction was significant, it was investigated further using four separate two-way ANOVAs. First, two-way ANOVAs





**Fig. 8.** Test group grand-averaged ERP waveforms elicited in the 900 ms interval following stimulus onset (0 ms) in response to standard (dashed lines), outer deviant (tW; gray lines), and central deviant (tB; black lines) blue color stimuli. The horizontal dashed line represents baseline (0  $\mu\text{V}$ ).

with Stimulus (collapsed across electrode sites within the relevant scalp region) and Color Region as factors were conducted separately for the test group and the control group. Second, separate two-way ANOVAs with Stimulus (collapsed across electrode sites

within the relevant scalp region) and Group as factors were conducted for the green color region and the blue color region. Significant two-way interactions (between Stimulus and Color Region; and Stimulus and Group, respectively) were followed up with a

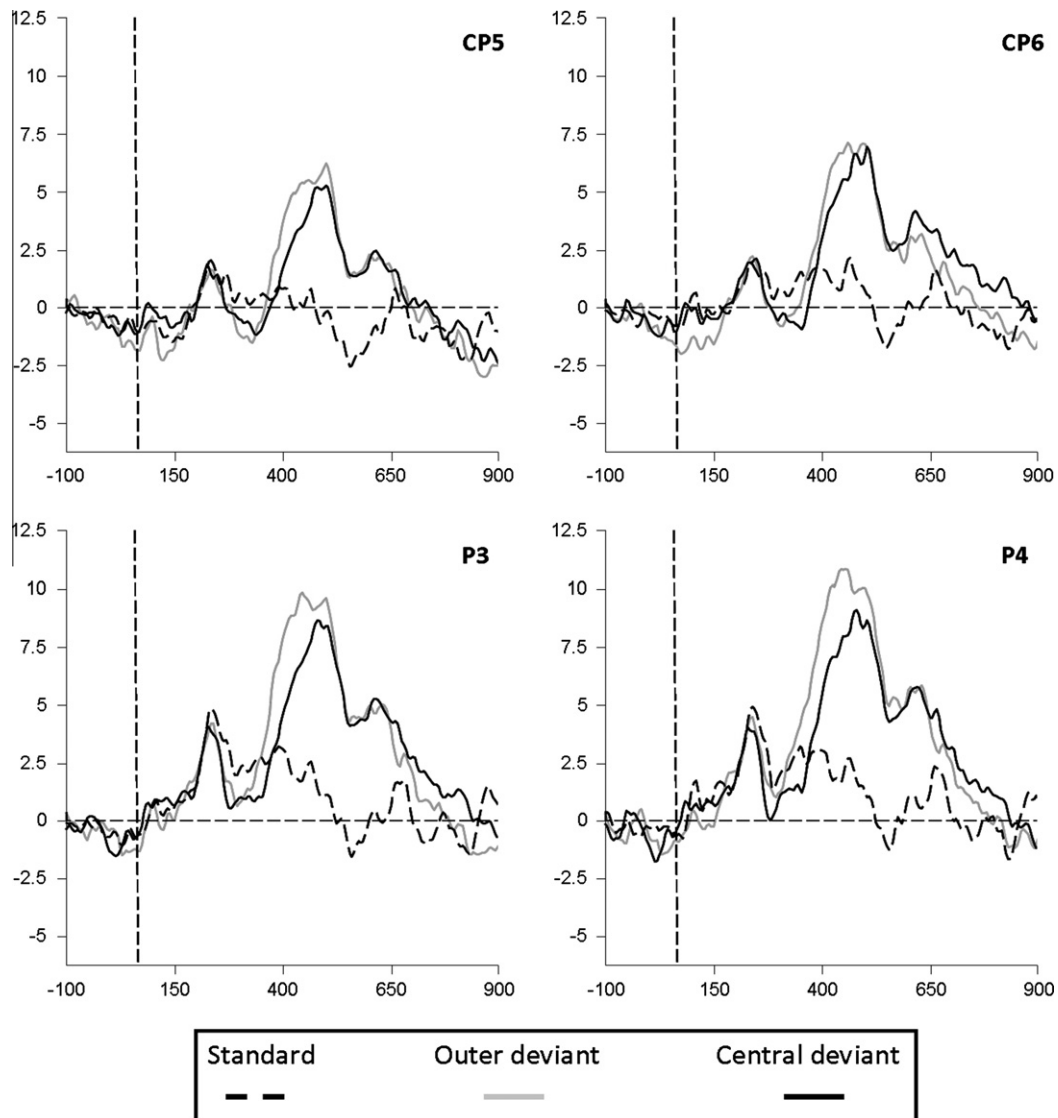


Fig. 8 (continued)

series of paired samples *t*-tests comparing the mean amplitude of outer deviants (tW) and central deviants (tB). Greenhouse-Geisser adjustments to the degrees of freedom were performed where appropriate, to correct for sphericity violations.

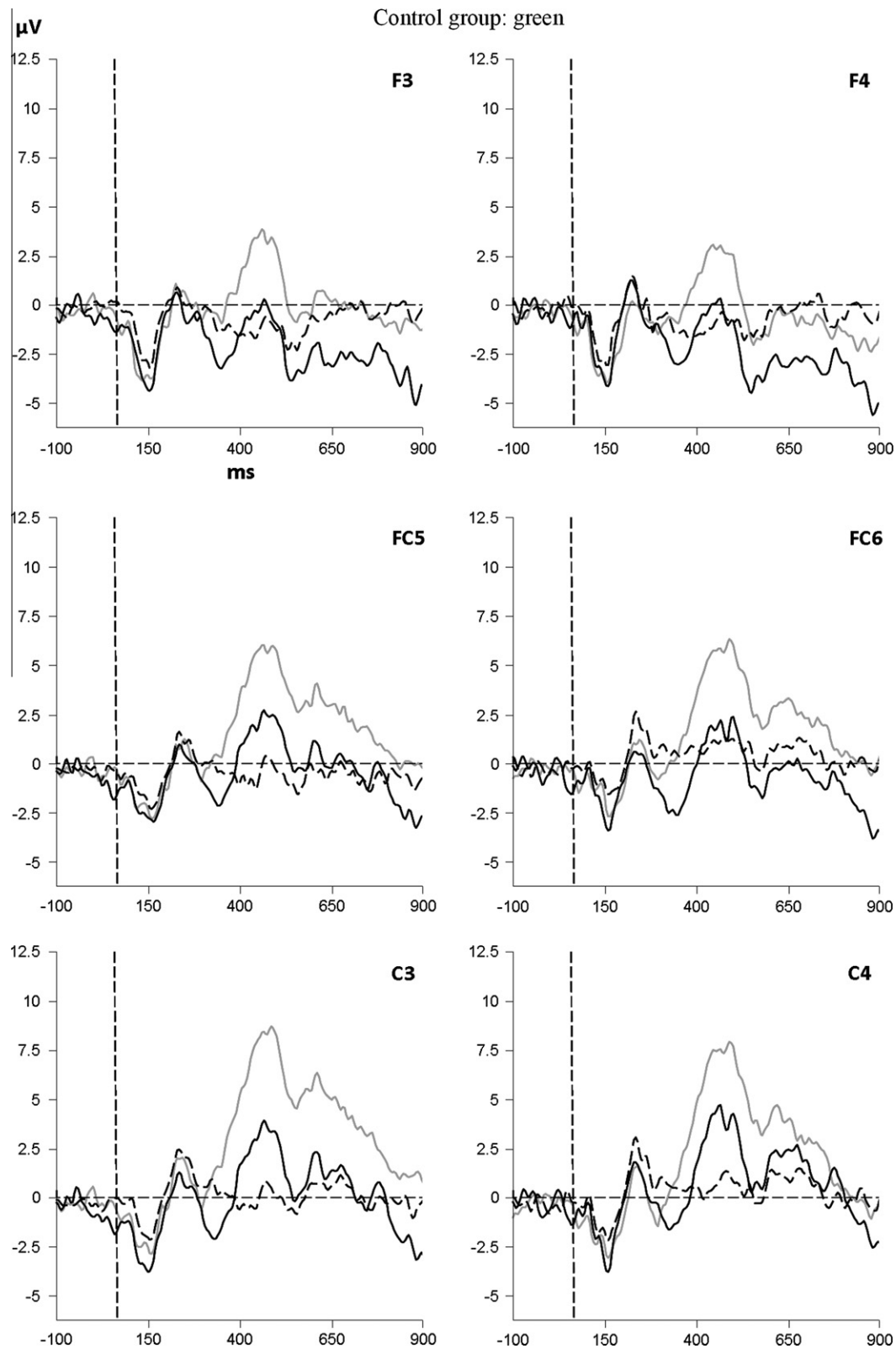
### 3.2.2. Mean amplitude

Figs. 7–10 show ERPs across a representative sample of electrode sites, elicited in response to standard (dashed lines), outer deviant (tW; gray lines) and central deviant (tB; black lines) stimuli. Fig. 7 (control group) and Fig. 8 (test group) show the waveforms for stimuli in the blue (no training) region of color space. Fig. 9 (control group) and Fig. 10 (test group) show the waveforms for stimuli in the green (training) region of color space. Upon examination of Figs. 7–10, it appears that there is no difference in mean amplitude for outer deviants (tW) and central deviants (tB) during the early and mid time ranges for all conditions. However, during the P3 time window there appear to be differences in mean amplitude between the two deviants. For the control group, blue outer deviants appear to elicit a greater P3 than blue central deviants across all scalp regions. This also appears to be the case for the test group, but to a lesser extent. Similarly, for the control group, green outer deviants elicit a greater P3 than green central

deviants. Most interestingly, however, for the test group (who have been trained on the green region of color space), it is the green central deviants (tB) that appear to elicit a greater P3 mean amplitude than the green outer deviants (tW). These impressions were supported by statistical analysis.

**3.2.2.1. P1, N1, P2, N2 time ranges.** Four-way ANOVA revealed that there was no significant interaction between Stimulus, Color Region and Group at all sites during the early and mid time ranges (all  $F_s < 3.35$ , all  $p_s > .079$ ).

**3.2.2.2. P3 time range; anterior sites.** For the mean amplitude elicited by outer deviants (tW) and central deviants (tB) in each condition see Fig. 11. Four-way ANOVA revealed that at anterior sites within the P3 time range there was a significant interaction between Stimulus, Color Region and Group ( $F(1,13) = 5.14$ ,  $p < .05$ ,  $\eta_p^2 = .17$ ). This crucial interaction was investigated further using a series of two-way ANOVAs. A two-way ANOVA with Stimulus (collapsed across electrode sites F3/4, F7/8, FC5/6) and Group as factors was conducted for the blue region of color space. However, this did not reveal any differences in the pattern of amplitude for blue outer deviants and central deviants for the control and test



**Fig. 9.** Control group grand-averaged ERP waveforms elicited in the 900 ms interval following stimulus onset (0 ms) in response to standard (dashed lines), outer deviant (tW; gray lines), and central deviant (tB; black lines) green color stimuli. The horizontal dashed line represents baseline (0 µV).

groups ( $F(1,26) = 1.34$ ,  $p = .257$ ,  $\eta_p^2 = .05$ ). The same two-way ANOVA was conducted for the green region of color space. Here, there was an interaction between Stimulus and Group ( $F(1,26) = 26.86$ ,  $p < .001$ ,  $\eta_p^2 = .51$ ). This was followed up using paired samples

$t$ -tests, which revealed that mean amplitude for outer deviants and central deviants differed for the control group ( $t(13) = 3.71$ ,  $p < .005$ ,  $d = 0.80$ ), with outer deviants eliciting significantly greater P3 mean amplitude. For the test group, outer deviants (tW) and the

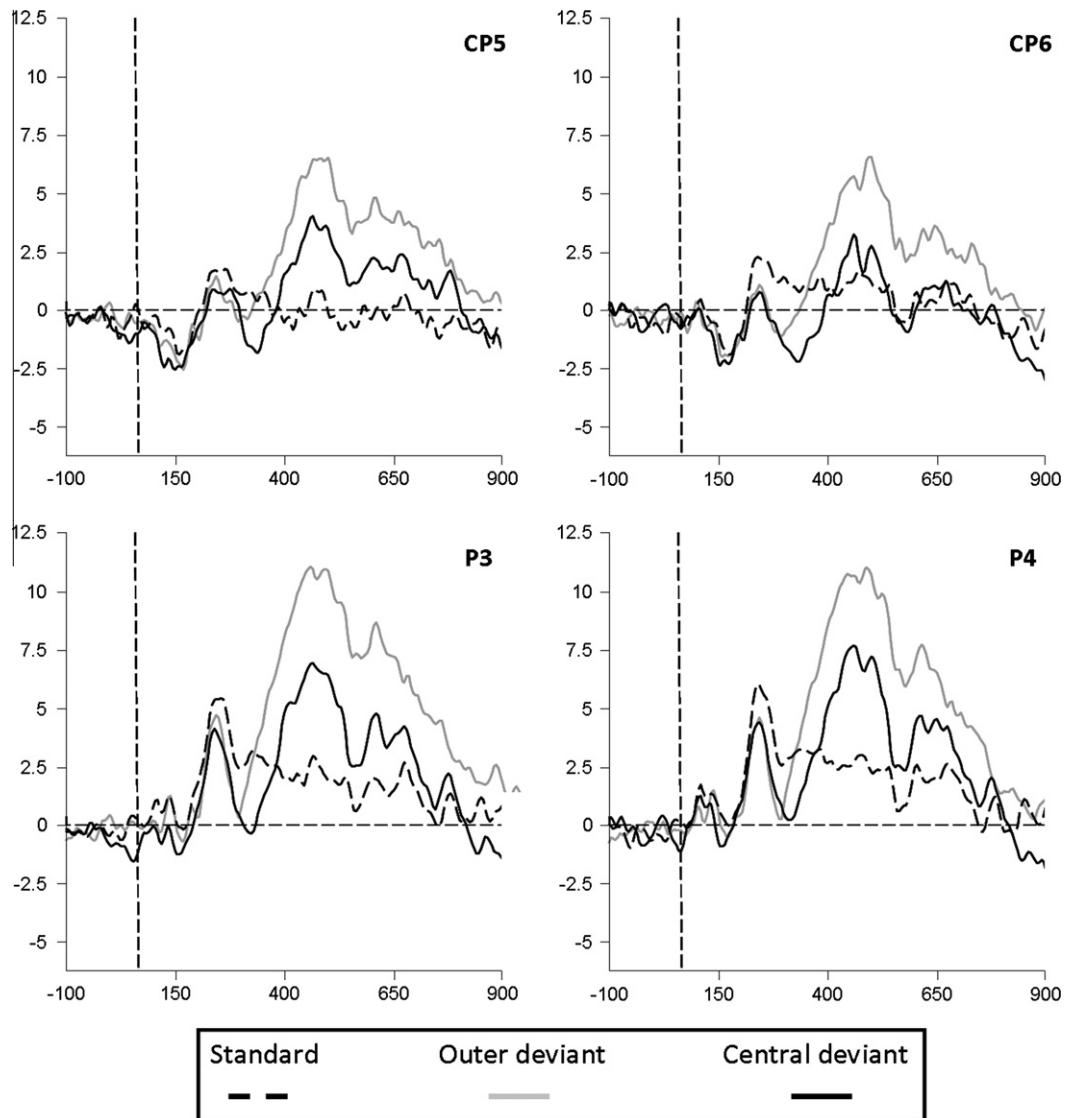


Fig. 9 (continued)

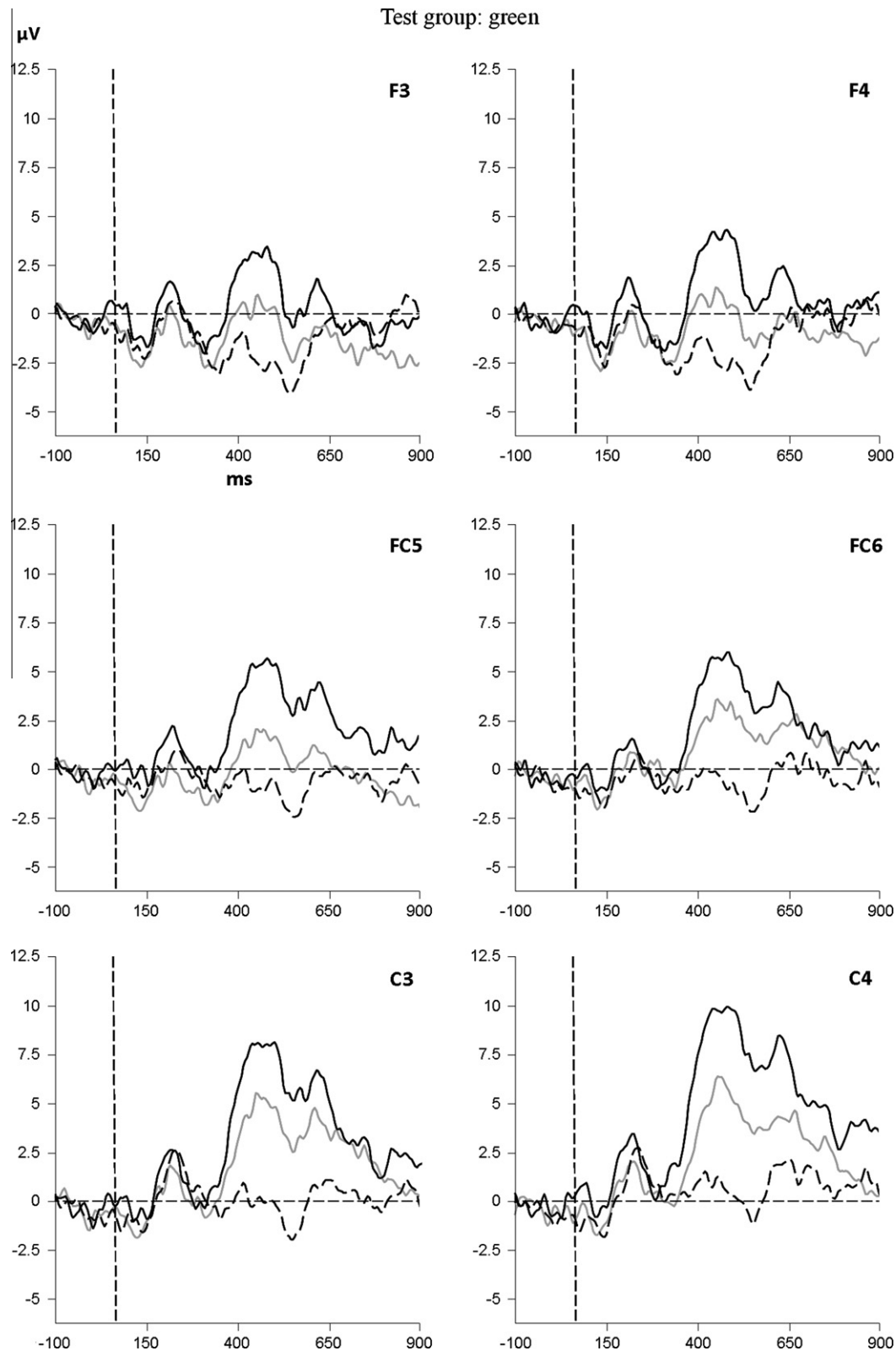
central deviants (tB) also differed ( $t(13) = 4.22$ ,  $p < .001$ ,  $d = 1.26$ ), but importantly central deviants (tB) elicited greater P3 mean amplitude than outer deviants (tW).

To further explore the interaction between Stimulus, Color Region and Group, a two-way ANOVA with Stimulus (collapsed across electrode sites F3/4, F7/8, FC5/6) and Color Region as factors was conducted on the control group data. This revealed no differences in the pattern of mean amplitude for outer deviants and central deviants across color regions for the control group ( $F(1, 13) = 1.76$ ,  $p = .208$ ,  $\eta_p^2 = .12$ ). The same two-way ANOVA was also conducted on the test group data. Importantly, there was a significant interaction between Stimulus and Color Region ( $F(1, 13) = 4.99$ ,  $p < .05$ ,  $\eta_p^2 = .28$ ), indicating that for the test group, the pattern of mean amplitude for outer deviants (tW) and central deviants (tB) differed for the green and blue color regions. Paired samples  $t$ -tests confirmed that there was no difference in the mean amplitude for outer deviants and central deviants in the blue region of color space, on which the test group had received no training. However, category effects were found in the green region of color space, on which the test group had been trained ( $t(13) = 4.22$ ,  $p = .001$ ,  $d = 1.26$ ), with central deviants (tB) eliciting a greater P3 mean amplitude than outer deviants (tW).

Finally, independent samples  $t$ -tests were performed to investigate the impact of training on the discrimination of outer deviants (tW) and central deviants (tB) in the green region of color space. Green outer deviant (tW) mean amplitude was compared for the control and the test group. There was no significant difference between the groups ( $t(13) = 1.41$ ,  $p = .171$ ,  $d = 0.54$ ), indicating that training did not significantly affect the processing of outer deviants (tW). Green central deviant (tB) mean amplitude was also compared for the two groups, revealing that training had a significant impact on the processing of central deviants (tB;  $t(13) = 2.26$ ,  $p < .05$ ,  $d = 0.85$ ).

**3.2.2.3. P3 time range; midline sites.** See Fig. 12 for the mean amplitude elicited by outer deviants (tW) and central deviants (tB) for the control and test groups in both color regions. An identical pattern of results to that found at anterior sites was revealed for midline sites within the P3 time range.

**3.2.2.4. P3 time range; posterior sites.** For the mean amplitude elicited by outer deviants (tW) and central deviants (tB) across all conditions see Fig. 13. The pattern of results found at anterior and midline sites was largely replicated over posterior regions.



**Fig. 10.** Test group grand-averaged ERP waveforms elicited in the 900 ms interval following stimulus onset (0 ms) in response to standard (tW; gray lines), and central deviant (tB; black lines) green color stimuli. The horizontal dashed line represents baseline (0  $\mu\text{V}$ ).

However, for the test group the difference between the untrained blue region and the trained green region was even greater at posterior sites ( $F(1, 13) = 20.06$ ,  $p = .001$ ,  $\eta_p^2 = .61$ ).

### 3.2.3. Peak latencies

It appears from inspection of the waveforms that there are no differences in the pattern of peak latencies for outer deviants



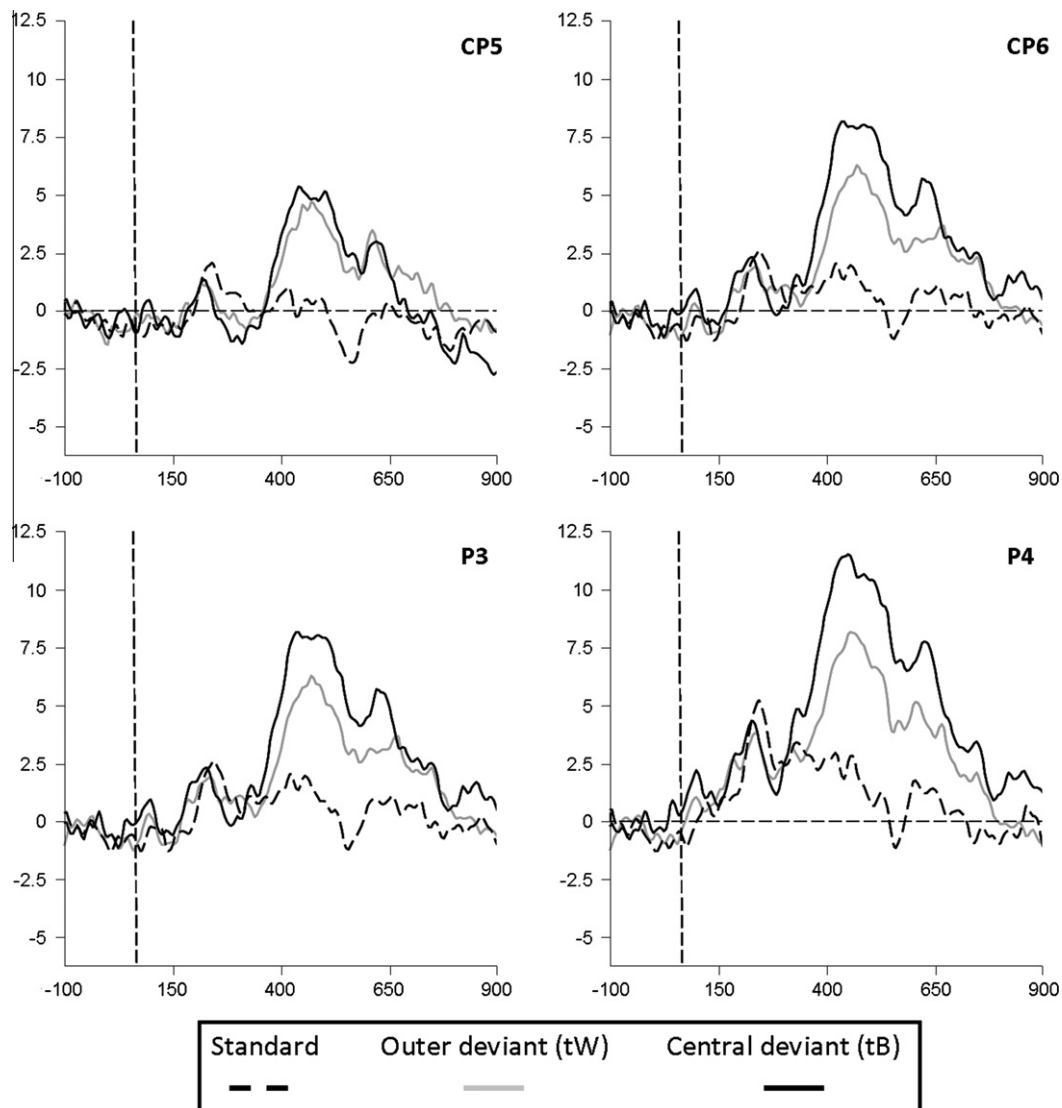


Fig. 10 (continued)

(tW) and central deviants (tB) across conditions and time windows. This observation was supported by statistical analysis. The crucial interaction between Stimulus, Color Region and Group approached significance for P1 at anterior areas ( $F(1,26) = 3.94$ ,  $p = .058$ ,  $\eta_p^2 = .13$ ) but was non-significant for each of the other component time ranges at all scalp regions (largest  $F = 2.73$ , smallest  $p = .111$ ).

### 3.3. Counting task

The mean performance for identifying the occurrence of oddball events was 97% ( $SD = 5$ ).

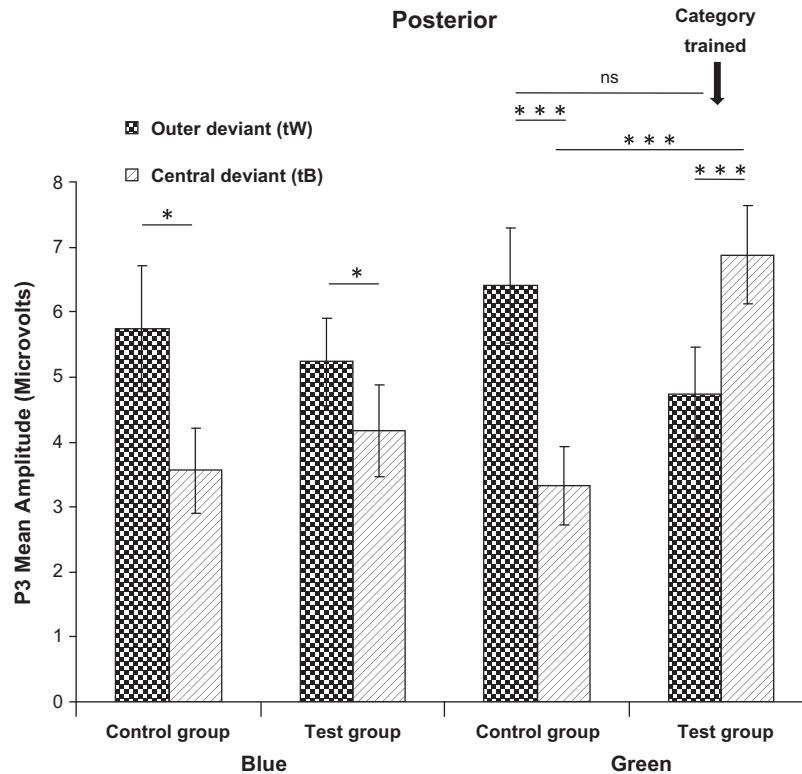
## 4. Discussion

Test group participants were trained to divide the green region of color space into two new categories and were found to be significantly more accurate at doing this on the last day of training compared to the first day. After the training phase, ERPs were recorded during a visual oddball task and were compared for the test group and a control group who had received no prior training. No

differences between the two groups were found during the early and mid time ranges (P1, N1, P2 and N2). However, differences were revealed within the P3 time range. Comparison of P3 amplitudes elicited in response to stimuli from the green region of color space showed significant effects of training. For the control group, green outer deviants evoked a greater P3 than green central deviants, as predicted. However, for the test group, who were trained on the green region of color space, this pattern was reversed: green central deviants that were from a different newly trained category to the standard evoked greater P3 mean amplitude than green outer deviants that were from the same newly trained category as the standard.

P3 amplitudes elicited in response to stimuli from the untrained blue region of color space were also compared for the test group and the control group. Participants in both groups received no training on the blue stimuli and so this comparison revealed whether there were any differences between the two groups that may have accounted for the differences outlined above. The same pattern of results was found for test and control group participants. For both groups, blue outer deviants elicited a greater P3 than blue central deviants – the same pattern elicited by green deviants for the control group. This suggests that the differences between the





**Fig. 13.** Posterior P3 mean amplitude for outer deviants (tW) and central deviants (tB) in the blue region of color space and the green region of color space, for both control and test groups. Significant differences are starred: \*\*\* $p < .005$ ; \* $p < .05$ . Error bars represent  $\pm 1SE$ .

perceptual categories compared to focal regions (e.g., Harnad, 1987; Kuhl, 1991; Lively & Pisoni, 1997; Özgen & Davies, 2002). The reversal of this pattern following training indicates the strength of the acquired category effects.

These acquired category effects were investigated further by exploring whether category training resulted in greater P3 amplitude for central deviants (tB; acquired distinctiveness—where between-category stimuli become more different), or reduced P3 amplitude for outer deviants (tW; acquired equivalence—where within-category stimuli become more similar). Green outer deviant (tW) P3 amplitude was compared for the test and the control group. No differences between the groups were revealed, with outer deviant (tW) amplitude being similar for the two groups, showing no acquired equivalence. Conversely, central deviant (tB) P3 amplitude for the test group and the control group was found to differ. Central deviant (tB) amplitude was significantly larger for the test group compared to the control group, showing acquired distinctiveness. This is consistent with the findings of Özgen and Davies (2002) and Goldstone (1994). This finding suggests that category learning enhances sensitivity around new category boundaries but does not necessarily reduce sensitivity within the new category. Therefore, this heightened sensitivity at new boundary regions appears to be a pervasive mechanism in category learning.

Further evidence for the processes involved in these acquired category effects comes from the time course of ERP waveforms. Acquired category effects were not found in early ERP components (P1, N1, P2 and N2). Inspection of the ERP waveforms indicates the absence of a clear P1 at anterior and midline sites, which may be surprising. However, the P1 is evident at posterior scalp locations, where it is typically maximal (see e.g., Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002). Category effects in P1 and N1 components have been found on established category

boundaries (see Holmes et al., 2009). As the current study used the same technique and task and employed a similar analysis with a comparable number of trials to the Holmes et al. study, it is unlikely that the current finding could be explained by anything other than the acquired nature of these category effects.

Although the current study found no category effects in early ERP components, category effects were present during a post-perceptual ERP component (P3). This suggests that acquired category effects can exist without the involvement of early perceptual processing and can be solely governed by cognitive mechanisms such as memory and language (e.g., McCarthy & Donchin, 1981). This study therefore provides further confirmation that category effects are not necessarily perceptual, and that the term 'categorical perception' which is traditionally used to describe such effects is misleading and should only be used where there is clear evidence for the involvement of perceptual processes. The absence of category effects in early perceptual ERP components following category training is in contrast to the presence of early perceptual category effects across the blue–green category boundary (e.g., Holmes et al., 2009). It is possible that early perceptual category effects for novel categories could be acquired following category training that is longer or more intensive than that undergone in the current study. Alternatively, the early perceptual category effects across category boundaries such as blue–green could depend on 'in-born tendencies' (e.g., Bornstein et al., 1976; Clifford et al., 2009; Franklin & Davies, 2004; Franklin, Drivonikou, Bevis, et al., 2008). Analysis of 110 of the world's color lexicons does suggest that there are universal constraints that lead to blue and green categories forming in a color lexicon (e.g., Kay & Regier, 2007; Regier, Kay & Khertarpal, 2007), and infants also respond categorically across the blue–green boundary (e.g., Clifford et al., 2009; Franklin & Davies, 2004; Franklin, Drivonikou, Bevis, et al., 2008). Therefore, categories such as blue and green may have a special status in perception relative to novel

categories that make arbitrary divisions of the color space. Further research is needed to investigate these issues.

The current study makes a number of important theoretical contributions to the literature. First, the study contributes to ongoing and extensive debate on the nature of color category effects. In particular, there has been recent controversy over whether such category effects for color are even categorical (Brown et al., 2011; Witzel & Gegenfurtner, 2011). For example, some have argued that supposed color category effects on tasks such as visual search, are better explained by cone-excitation or inequalities in the color metric used to equate same- and different-category color pairs (Brown et al., 2011; Witzel & Gegenfurtner, 2011). This may well account for many of the so-called 'category effects' in the literature. However, here we provide a clear example of a category effect that cannot be explained in terms of cone-excitation or inequalities in color metric. The stimuli were the same before and after training, and were the same for the trained and the non-trained group, and therefore the category effect can clearly be attributed to the newly trained categories. Therefore, the study verifies that newly learned color categories can affect the detection of color differences.

Second, the current study provides clear evidence that the influence of color categories on the detection of color differences is not necessarily due to changes in early perceptual processes, and that categories can affect higher-order processes in isolation. Although previous studies have provided evidence that categories and concepts penetrate perception (e.g., Lupyan et al., 2010; Notman et al., 2005; Reber et al., 1998; Tanaka & Curran, 2001), it is clear from the current study that this is not always the case. Therefore, the findings support the suggestion that the term 'categorical perception' is a misleading one, and suggest that the term should only be used when a clear perceptual locus for the effect is known.

Third, the current study illustrates how the ERP method can provide precision on the extent to which categories penetrate perception and the extent to which learning new categories leads to perceptual change. Some have argued, on the basis of behavioral studies, that categories and concepts penetrate perception (e.g., Lupyan et al., 2010). The current study highlights how the ERP method could be useful in confirming whether category effects found in previous behavioral studies really are located in early perception. Categories and concepts may well affect the way in which we 'see' the world on many different levels. Further research is needed to clarify the conditions under which categories actually do change our perceptual representation of the world.

## Acknowledgments

This research was supported by a departmental PhD bursary to Alexandra Clifford from the Department of Psychology, University of Surrey. We thank Simon Mathews for assistance with computer programming.

## References

- Boroditsky, L. (2003). Linguistic relativity. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 917–921). London, UK: MacMillan Press.
- Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorisation in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 115–129.
- Bornstein, M. H., & Korda, N. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research*, 46, 207–222.
- Brown, A. M., Lindsey, D. T., & Guckes, K. M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision*, 11(12), 1–21.
- Clifford, A., Franklin, A., Davies, I. R. L., & Holmes, A. (2009). Electrophysiological markers of categorical perception of color in 7-month old infants. *Brain and Cognition*, 71, 165–172.
- Clifford, A., Holmes, A., Davies, I. R. L., & Franklin, A. (2010). Color categories affect pre-attentive color perception. *Biological Psychology*, 85(2), 275–282.
- Dalvit, S., & Eimer, M. (2011). Memory-driven attentional capture is modulated by temporal task demands. *Visual Cognition*, 19, 145–153.
- Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, 15, 95–111.
- Drivonikou, G. V., Clifford, A., Franklin, A., & Davies, I. R. L. (2011). Category training affects colour discrimination but only in the right visual field. In C. P. Biggam, C. Hough, C. J. Kay, & D. R. C. Simmons (Eds.), *Progress in colour studies: New directions in colour studies*. Amsterdam, NL: John Benjamins.
- Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences*, 104(3), 1097–1102.
- Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, 22, 349–377.
- Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences*, 105, 3221–3225.
- Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences*, 105(47), 18221–18225.
- Franklin, A., Pilling, M., & Davies, I. R. L. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Child Psychology*, 91, 227–248.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, 103(2), 489–494.
- Goldstone, R. L. (1994). Influences of categorisation on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2), 178–200.
- Goldstone, R. L., Gerganov, A., Landy, D., & Roberts, M. E. (2008). Learning to see and conceive. In L. Tommasi, M. Peterson, & L. Nadel (Eds.), *The new cognitive sciences (part of the Vienna series in theoretical biology)* (pp. 163–188). Cambridge, MA: MIT Press.
- Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 535–565). Cambridge: Cambridge University Press.
- Holmes, A., Franklin, A., Clifford, A., & Davies, I. R. L. (2009). Neurophysiological evidence for categorical perception of color. *Brain and Cognition*, 69, 426–434.
- Ishihara, S. (1987). *Ishihara test for colour-blindness*. Tokyo: Kanehara & Co. Ltd.
- Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371–375.
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, 86, 65–79.
- Kay, P., & Regier, T. (2007). Color naming universals: The case of Berlinmo. *Cognition*, 102(2007), 289–298.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93–107.
- Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1665–1679.
- Lupyan, G., Thompson-Schill, S. L., & Swingle, D. (2010). Conceptual penetration of visual processing. *Psychological Science*, 21(5), 682–691.
- McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, 211, 77–80.
- Mishra, J., Martinez, A., Schroeder, C. E., & Hillyard, S. A. (2012). *NeuroImage*, 59, 1968–1978.
- Newhall, S. M., Nickerson, D., & Judd, B. D. (1943). Final report of the O.S.A. subcommittee on the spacing of the Munsell colours. *Journal of the Optical Society of America*, 33(7), 385–418.
- Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition*, 95(2), B1–B14.
- Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, 131(4), 477–493.
- Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *International Journal of Medical Sciences*, 2, 147–154.
- Polich, J. (1999). P300 in clinical applications. In E. Niedermeyer, F. L. d. Silva (Eds.), *Electroencephalography: Basic Principles, clinical applications and related fields* (4th ed., pp. 1073–1091). Baltimore-Munich: Urban & Schwarzenberg.
- Pylshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–365.
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning and Memory*, 5, 420–428.
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 104, 1436–1441.
- Roberson, D., Pak, H., & Hanley, R. J. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, 107(2), 752–762.
- Roberson, D., Hanley, R. J., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, 112, 482–487.

- Rugg, M. D., & Coles, M. G. H. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford: Oxford University Press.
- Siok, W. T., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luke, K.-K., et al. (2009). Language regions of the brain are operative in colour perception. *Proceedings of the National Academy of Sciences*, 106(20), 8140–8145.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, 12(1), 43–47.
- Taylor, M. J. (2002). Non-spatial attentional effects on P1. *Clinical Neurophysiology*, 113, 1903–1908.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104, 7780–7785.
- Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, 11(12), 1534–7362.
- Wyszecki, G., & Styles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae* (2nd ed.). New York: Wiley.
- Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences*, 107(22), 9974–9978.